

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-061098

(43)Date of publication of application : 28.02.2003

(51)Int.Cl.

H04N 7/32  
G09B 21/00  
G10L 15/00  
H04N 5/272  
H04N 5/76  
H04N 7/08  
H04N 7/081

(21)Application number : 2001-250392

(71)Applicant : CANON INC

(22)Date of filing : 21.08.2001

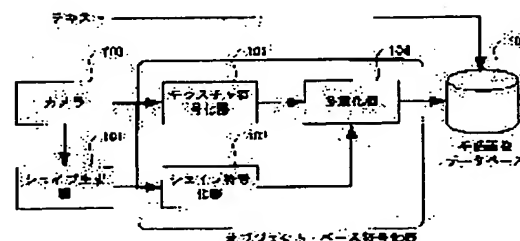
(72)Inventor : OHIRA TADASHI

(54) IMAGE PROCESSOR, IMAGE PROCESSING METHOD, RECORDING MEDIUM AND PROGRAM

(57)Abstract:

**PROBLEM TO BE SOLVED:** To provide an image processor that composites a sign language image together principal moving images and audio signals, so as to fully provide hard of hearing persons with a natural and easily understandable video services.

**SOLUTION:** The image processor is provided for encoding a moving image on an object basis to generate a database for sign language images. The processor is provided with a moving picture entry means for receiving a moving image of people making communication in sign language, an object extracting means that extracts only the people making communication in sign language as objects for capturing moving images of shape information, a natural image encoding means that encodes the moving picture of the people making communication in sign language, a shape information image encoding means that encodes the moving image of the information on the shape, a multiplexer means that multiplexes the natural image after encoding with the shape information image, a text input means that receives the meaning of the sign language image as text data, and a database storage means that stores the database comprising the multiplexed image and the text data.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of

Searching PAJ

rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2003-61098

(P2003-61098A)

(43)公開日 平成15年2月28日(2003.2.28)

(51)Int.Cl.	識別記号	F I	特許庁(参考)
H 0 4 N	7/32	G 0 9 B 21/00	F 5 C 0 2 3
G 0 9 B	21/00	H 0 4 N 5/272	5 C 0 5 2
G 1 0 L	15/00	5/76	B 5 C 0 5 9
H 0 4 N	5/272	7/137	Z 5 C 0 6 3
5/76		7/08	Z 5 D 0 1 5

審査請求 未請求 請求項の数18 O L (全 10 頁) 最終頁に続く

(21)出願番号 特願2001-250392(P2001-250392)

(22)出願日 平成13年8月21日(2001.8.21)

(71)出願人 000001007

キヤノン株式会社

東京都大田区下丸子3丁目30番2号

(72)発明者 大平 正

東京都大田区下丸子3丁目30番2号 キヤ

ノン株式会社内

(74)代理人 100090273

弁理士 園分 季悦

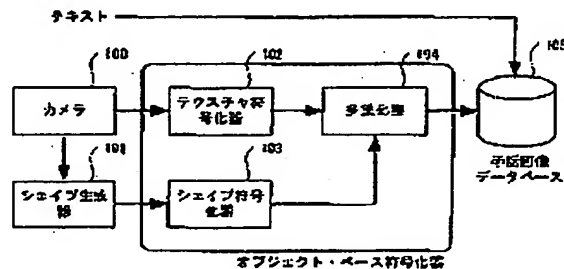
最終頁に続く

(54)【発明の名称】 画像処理装置、画像処理方法、記録媒体及びプログラム

(57)【要約】

【課題】 主たる動画像及びオーディオ信号に手話画像を合成することで、聴覚障害者にとって自然でわかりやすい映像サービスを充足させることを課題とする。

【解決手段】 動画像のオブジェクト・ベース符号化を行い、手話画像のデータベースを作成する画像処理装置であって、手話者の動画像を入力する動画像入力手段と、手話者のみをオブジェクトとして抽出し形状情報の動画像を獲得するオブジェクト抽出手段と、手話者の動画像を符号化する自然画像符号化手段と、形状情報の動画像を符号化する形状情報画像符号化手段と、符号化後の自然画像と形状情報画像とを多変化する多変化手段と、手話画像の意味をテキストデータとして入力するテキスト入力手段と、多変化した画像とテキストデータをデータベース化して蓄積するデータベース蓄積手段とを備えた画像処理装置が提供される。



## 【特許請求の範囲】

【請求項1】 手話画像のデータベースを作成するデータベース作成手段と、

主たる動画画像を符号化して前記データベース内の手話画像と多重化する多重化手段と、

前記多重化された画像から主たる動画画像と手話画像を復号し、合成出力する手段とを備えた画像処理装置、

【請求項2】 動画画像のオブジェクト・ベース符号化を行い、手話画像のデータベースを作成する画像処理装置であって、

手話者の動画画像を入力する動画画像入力手段と、

手話者のみをオブジェクトとして抽出し形状情報の動画画像を獲得するオブジェクト抽出手段と、

前記手話者の動画画像を符号化する自然画像符号化手段と、

前記形状情報の動画画像を符号化する形状情報画像符号化手段と、

前記符号化後の自然画像と形状情報画像とを多重化する多重化手段と、

前記手話画像の意味をテキストデータとして入力するテキスト入力手段と、

前記多重化した画像と前記テキストデータをデータベース化して蓄積するデータベース蓄積手段とを備えた画像処理装置、

【請求項3】 主たる動画画像信号とオーディオ信号を符号化する符号化手段と、

前記オーディオ信号から音声の情報を抽出し認識する音声認識手段と、

前記認識した音声データを利用し手話画像のデータベースから所望の手話画像を検索する手話画像検索手段と、

前記符号化した主たる動画画像信号とオーディオ信号と前記手話画像を多重化して出力する多重化手段とを備えた画像処理装置、

【請求項4】 入力信号を、主たる動画画像信号及びオーディオ信号の組みと、手話画像とに分配する第1の分配手段と、

前記主たる動画画像信号及びオーディオ信号を復号する第1の復号手段と、

前記手話画像から、手話者の自然画像と手話者の形状情報画像とに分配する第2の分配手段と、

前記手話者の自然画像を復号する第2の復号手段と、

前記手話者の形状情報画像を復号する第3の復号手段と、

前記主たる動画画像信号と前記手話者の自然画像と形状情報画像から、主たる動画画像と手話画像を合成する合成手段とを備えた画像処理装置、

【請求項5】 前記合成手段は、

前記主たる動画画像と前記手話画像とを合成することを指示する指示信号を入力する入力手段と、

前記指示信号を受けて、指示信号が合成指示する場合、

主たる動画画像と手話画像を合成し、合成指示しない場合、主たる動画画像と手話画像を合成しないで主たる動画画像を出力する切り替え手段とを備えた請求項4記載の画像処理装置、

【請求項6】 前記切り替え手段は、指示信号が合成指示する場合、主たる動画画像と手話画像を合成して出力し、合成指示しない場合、主たる動画画像とオーディオ信号を出力する請求項5記載の画像処理装置、

【請求項7】 手話画像のデータベースを作成するデータベース作成ステップと、

主たる動画画像を符号化して前記データベース内の手話画像と多重化する多重化ステップと、

前記多重化された画像から主たる動画画像と手話画像を復号し、合成出力するステップとを備えた画像処理方法、

【請求項8】 動画画像のオブジェクト・ベース符号化を行い、手話画像のデータベースを作成する画像処理方法であって、

手話者の動画画像を入力する動画画像入力ステップと、

手話者のみをオブジェクトとして抽出し形状情報の動画画像を獲得するオブジェクト抽出ステップと、

前記手話者の動画画像を符号化する自然画像符号化ステップと、

前記形状情報の動画画像を符号化する形状情報画像符号化ステップと、

前記符号化後の自然画像と形状情報画像とを多重化する多重化ステップと、

前記手話画像の意味をテキストデータとして入力するテキスト入力ステップと、

前記多重化した画像と前記テキストデータをデータベース化して蓄積するデータベース蓄積ステップとを備えた画像処理方法、

【請求項9】 主たる動画画像信号とオーディオ信号を符号化する符号化ステップと、

前記オーディオ信号から音声の情報を抽出し認識する音声認識ステップと、

前記認識した音声データを利用し手話画像のデータベースから所望の手話画像を検索する手話画像検索ステップと、

前記符号化した主たる動画画像信号とオーディオ信号と前記手話画像を多重化して出力する多重化ステップとを備えた画像処理方法、

【請求項10】 入力信号を、主たる動画画像信号及びオーディオ信号の組みと、手話画像とに分配する第1の分配ステップと、

前記主たる動画画像信号及びオーディオ信号を復号する第1の復号ステップと、

前記手話画像から、手話者の自然画像と手話者の形状情報画像とに分配する第2の分配ステップと、

前記手話者の自然画像を復号する第2の復号ステップと、

前記手話者の形状情報画像を復号する第3の復号ステップと、

前記主たる動画画像信号と前記手話者の自然画像と形状情報画像から、主たる動画画像と手話画像を合成する合成手段とを備えた画像処理装置、

前記合成手段は、

前記主たる動画画像と前記手話画像とを合成することを指示する指示信号を入力する入力手段と、

前記指示信号を受けて、指示信号が合成指示する場合、

前記手話者の形状情報画像を復号する第3の復号ステップと、

前記主たる動画画像信号と前記手話者の自然画像と形状情報画像から、主たる動画画像と手話画像を合成する合成ステップとを備えた画像処理方法。

【請求項11】 手話画像のデータベースを作成するデータベース作成手順と、

主たる動画画像を符号化して前記データベース内の手話画像と多重化する多重化手順と、

前記多重化された画像から主たる動画画像と手話画像を復号し、合成出力する手順とをコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項12】 動画画像のオブジェクト・ベース符号化を行い、手話画像のデータベースを作成するプログラムを記録した記録媒体であって、

手話者の動画画像を入力する動画画像入力手順と、

手話者のみをオブジェクトとして抽出し形状情報の動画画像を獲得するオブジェクト抽出手順と、

前記手話者の動画画像を符号化する自然画像符号化手順と、

前記形状情報の動画画像を符号化する形状情報画像符号化手順と、

前記符号化後の自然画像と形状情報画像とを多重化する多重化手順と、

前記手話画像の意味をテキストデータとして入力するテキスト入力手順と、

前記多重化した画像と前記テキストデータをデータベース化して蓄積するデータベース蓄積手順とをコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項13】 主たる動画画像信号とオーディオ信号を符号化する符号化手順と、

前記オーディオ信号から音声の情報を抽出し認識する音声認識手順と、

前記認識した音声データを利用し手話画像のデータベースから所望の手話画像を検索する手話画像検索手順と、

前記符号化した主たる動画画像信号とオーディオ信号と前記手話画像を多重化して出力する多重化手順とをコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項14】 入力信号を、主たる動画画像信号及びオーディオ信号の組合と、手話画像とに分配する第1の分配手順と、

前記主たる動画画像信号及びオーディオ信号を復号する第1の復号手順と、

前記手話画像から、手話者の自然画像と手話者の形状情報画像とに分配する第2の分配手順と、

前記手話者の自然画像を復号する第2の復号手順と、

前記手話者の形状情報画像を復号する第3の復号手順

と、

前記主たる動画画像信号と前記手話者の自然画像と形状情報画像から、主たる動画画像と手話画像を合成する合成手順とをコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項15】 手話画像のデータベースを作成するデータベース作成手順と、

主たる動画画像を符号化して前記データベース内の手話画像と多重化する多重化手順と、

前記多重化された画像から主たる動画画像と手話画像を復号し、合成出力する手順とをコンピュータに実行させるためのプログラム。

【請求項16】 動画画像のオブジェクト・ベース符号化を行い、手話画像のデータベースを作成するプログラムであって、

手話者の動画画像を入力する動画画像入力手順と、

手話者のみをオブジェクトとして抽出し形状情報の動画画像を獲得するオブジェクト抽出手順と、

前記手話者の動画画像を符号化する自然画像符号化手順

と、

前記形状情報の動画画像を符号化する形状情報画像符号化手順と、

前記符号化後の自然画像と形状情報画像とを多重化する多重化手順と、

前記手話画像の意味をテキストデータとして入力するテキスト入力手順と、

前記多重化した画像と前記テキストデータをデータベース化して蓄積するデータベース蓄積手順とをコンピュータに実行させるためのプログラム。

【請求項17】 主たる動画画像信号とオーディオ信号を符号化する符号化手順と、

前記オーディオ信号から音声の情報を抽出し認識する音声認識手順と、

前記認識した音声データを利用し手話画像のデータベースから所望の手話画像を検索する手話画像検索手順と、

前記符号化した主たる動画画像信号とオーディオ信号と前記手話画像を多重化して出力する多重化手順とをコンピュータに実行させるためのプログラム。

【請求項18】 入力信号を、主たる動画画像信号及びオーディオ信号の組合と、手話画像とに分配する第1の分配手順と、

前記主たる動画画像信号及びオーディオ信号を復号する第1の復号手順と、

前記手話画像から、手話者の自然画像と手話者の形状情報画像とに分配する第2の分配手順と、

前記手話者の自然画像を復号する第2の復号手順と、

前記手話者の形状情報画像を復号する第3の復号手順

と、

前記主たる動画画像信号と前記手話者の自然画像と形状情報画像から、主たる動画画像と手話画像を合成する合成手

順とをコンピュータに実行させるためのプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、聴覚者から聴覚障害者への通訳を行うシステムに関するものであり、特に、画像と音声を持つコンテンツに対して手話画像を生成、多重化し、必要に応じて合成表示する技術に関するものである。

【0002】

【従来の技術】手話は、手の位置、向き、移動方向、移動速度、表情等の手段を使って情報を相手に伝える、聴覚障害者のための言語で、音声を中心として発達してきた聴覚者の用いている自然言語とは異なった体系を持っている。したがって、聴覚障害者が聴覚者と会話を行う際には音声言語の体系に属する自然言語を用いて筆談や口話を行うよりも、手話で会話を行った方が楽である上に伝達速度も速い。そのため、自然言語と手話を混在させたシステムが望まれている。

【0003】従来の手話を生成する方法としては、CGアニメーションを用いるものや単語単位で撮影した画像をそのままつないで表示する方法がある。

【0004】CGアニメーションを用いる方法では、単純なものでは細部の情報がわかりにくく手話で最も重要な手の動きがわかりにくくなる。対して複雑なアニメーションを作ろうとすると高度な開発環境を必要とする。

【0005】一方で、コンピュータ・放送・通信など多くの領域で利用できる、汎用的な次世代マルチメディア符号化規格として、ISO (International Organization for Standardization: 国際標準化機構) やITU

(International Telecommunication Union: 国際電気通信連合) によって、MPEG-4の国際標準化作業が進められている。MPEG-4の大きな特徴として、オブジェクト・ベース符号化を行う機能がある。

【0006】オブジェクト・ベース符号化とは、MPEG (Moving Picture Expert Group) -1やMPEG-2で採用されているような長方形の画像全体を符号化する方法ではなく、予め何等かの方法で生成された形状情報の画像によって、画像の中の切り出された人物やその他の物体、つまり画像のオブジェクト毎に符号化を行う方法である。以降ではこの形状情報画像と区別するため、一般的に処理対象とする画像を自然画像と称す。

【0007】形状情報画像とは、符号化対象となっている自然画像と全く同じ縦横の画素数を持ち、オブジェクトの形状を表す一値の画像である。形状情報画像には各画素が1bitで表されるバイナリ・アルファ・プレーンと、各画素が2bit以上で表されるグレイスケール・アルファ・プレーンがある。バイナリ・アルファ・プレーンは道筋、画素の値が“1”の領域はオブジェクト領域、“0”の領域はオブジェクト外の領域を表す。グレイスケール・アルファ・プレーンは、画素の値が“1”以

上255以下の領域はオブジェクト領域、“0”の領域はオブジェクト外の領域を表す。

【0008】尚、その他のMPEG-4の詳細内容については、ISO/IECによる国際標準の文書に委ねることとする。

【0009】

【発明が解決しようとする課題】手話での会話が円滑に行なわれるためには、手話者の画像が明瞭である必要があり、また手話者の手の形が正しく認識されるためには、画像は立体的に表示される必要がある。

【0010】従来から、手話を撮影した画像を単純に合成する方法では、画像は立体的であるが記憶すべき情報量が多い。手話単語を登録する手話画像の辞書には、一手話単語あたり最低60枚前後のカラー画像を2000単語以上記憶しなければならないことから、一手話単語あたりの記憶すべき情報量が多いため多くの記憶容量を必要とする。

【0011】本発明の目的は、オブジェクト・ベース符号化装置及びその方法を用いて、任意形状の手話画像のデータベースを効率的に構築する手段、及び主たるAVストリーム（全画面の動画像と音声）に手話画像を合成することで、聴覚障害者にとって自然でわかりやすい映像サービスを充足させる手段を提供することである。

【0012】

【課題を解決するための手段】本発明の一観点によれば、手話画像のデータベースを作成するデータベース作成手段と、主たる動画像を符号化して前記データベース内の手話画像と多重化する多重化手段と、前記多重化された画像から主たる動画像と手話画像を復号し、合成出力する手段とを備えた画像処理装置が提供される。

【0013】本発明の他の観点によれば、動画像のオブジェクト・ベース符号化を行い、手話画像のデータベースを作成する画像処理装置であって、手話者の動画像を入力する動画像入力手段と、手話者のみをオブジェクトとして抽出し形状情報の動画像を獲得するオブジェクト抽出手段と、前記手話者の動画像を符号化する自然画像符号化手段と、前記形状情報の動画像を符号化する形状情報画像符号化手段と、前記符号化後の自然画像と形状情報画像とを多重化する多重化手段と、前記手話画像の意味をテキストデータとして入力するテキスト入力手段と、前記多重化した画像と前記テキストデータをデータベース化して蓄積するデータベース蓄積手段とを備えた画像処理装置が提供される。

【0014】本発明のさらに他の観点によれば、主たる動画像信号とオーディオ信号を符号化する符号化手段と、前記オーディオ信号から音声の情報を抽出し認識する音声認識手段と、前記認識した音声データを利用し手話画像のデータベースから所望の手話画像を検索する手話画像検索手段と、前記符号化した主たる動画像信号とオーディオ信号と前記手話画像を多重化して出力する多

重化手段とを備えた画像処理装置が提供される。

【0015】本発明のさらに他の観点によれば、入力信号を、主たる動画画像信号及びオーディオ信号の組みと、手話画像とに分配する第1の分配手段と、前記主たる動画画像信号及びオーディオ信号を復号する第1の復号手段と、前記手話画像から、手話者の自然画像と手話者の形状情報画像とに分配する第2の分配手段と、前記手話者の自然画像を復号する第2の復号手段と、前記手話者の形状情報画像を復号する第3の復号手段と、前記主たる動画画像信号と前記手話者の自然画像と形状情報画像から、主たる動画画像と手話画像を合成する合成手段とを備えた画像処理装置が提供される。

【0016】本発明のさらに他の観点によれば、手話画像のデータベースを作成するデータベース作成ステップと、主たる動画画像を符号化して前記データベース内の手話画像と多重化する多重化ステップと、前記多重化された画像から主たる動画画像と手話画像を復号し、合成出力するステップとを備えた画像処理方法が提供される。

【0017】本発明のさらに他の観点によれば、動画画像のオブジェクト・ベース符号化を行い、手話画像のデータベースを作成する画像処理方法であって、手話者の動画画像を入力する動画画像入力ステップと、手話者のみをオブジェクトとして抽出し形状情報の動画画像を獲得するオブジェクト抽出ステップと、前記手話者の動画画像を符号化する自然画像符号化ステップと、前記形状情報の動画画像を符号化する形状情報画像符号化ステップと、前記符号化後の自然画像と形状情報画像とを多重化する多重化ステップと、前記手話画像の意味をテキストデータとして入力するテキスト入力ステップと、前記多重化した画像と前記テキストデータをデータベース化して蓄積するデータベース蓄積ステップとを備えた画像処理方法が提供される。

【0018】本発明のさらに他の観点によれば、主たる動画画像信号とオーディオ信号を符号化する符号化ステップと、前記オーディオ信号から音声の情報を抽出し認識する音声認識ステップと、前記認識した音声データを利用し手話画像のデータベースから所望の手話画像を検索する手話画像検索ステップと、前記符号化した主たる動画画像信号とオーディオ信号と前記手話画像を多重化して出力する多重化ステップとを備えた画像処理方法が提供される。

【0019】本発明のさらに他の観点によれば、入力信号を、主たる動画画像信号及びオーディオ信号の組みと、手話画像とに分配する第1の分配ステップと、前記主たる動画画像信号及びオーディオ信号を復号する第1の復号ステップと、前記手話画像から、手話者の自然画像と手話者の形状情報画像とに分配する第2の分配ステップと、前記手話者の自然画像を復号する第2の復号ステップと、前記手話者の形状情報画像を復号する第3の復号ステップと、前記主たる動画画像信号と前記手話者の自然

画像と形状情報画像から、主たる動画画像と手話画像を合成する合成ステップとを備えた画像処理方法が提供される。

【0020】本発明のさらに他の観点によれば、手話画像のデータベースを作成するデータベース作成手順と、主たる動画画像を符号化して前記データベース内の手話画像と多重化する多重化手順と、前記多重化された画像から主たる動画画像と手話画像を復号し、合成出力する手順とをコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体が提供される。

【0021】本発明のさらに他の観点によれば、動画画像のオブジェクト・ベース符号化を行い、手話画像のデータベースを作成するプログラムを記録した記録媒体であって、手話者の動画画像を入力する動画画像入力手順と、手話者のみをオブジェクトとして抽出し形状情報の動画画像を獲得するオブジェクト抽出手順と、前記手話者の動画画像を符号化する自然画像符号化手順と、前記形状情報の動画画像を符号化する形状情報画像符号化手順と、前記符号化後の自然画像と形状情報画像とを多重化する多重化手順と、前記手話画像の意味をテキストデータとして入力するテキスト入力手順と、前記多重化した画像と前記テキストデータをデータベース化して蓄積するデータベース蓄積手順とをコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体が提供される。

【0022】本発明のさらに他の観点によれば、主たる動画画像信号とオーディオ信号を符号化する符号化手順と、前記オーディオ信号から音声の情報を抽出し認識する音声認識手順と、前記認識した音声データを利用し手話画像のデータベースから所望の手話画像を検索する手話画像検索手順と、前記符号化した主たる動画画像信号とオーディオ信号と前記手話画像を多重化して出力する多重化手順とをコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体が提供される。

【0023】本発明のさらに他の観点によれば、入力信号を、主たる動画画像信号及びオーディオ信号の組みと、手話画像とに分配する第1の分配手順と、前記主たる動画画像信号及びオーディオ信号を復号する第1の復号手順と、前記手話画像から、手話者の自然画像と手話者の形状情報画像とに分配する第2の分配手順と、前記手話者の自然画像を復号する第2の復号手順と、前記手話者の形状情報画像を復号する第3の復号手順と、前記主たる動画画像信号と前記手話者の自然画像と形状情報画像から、主たる動画画像と手話画像を合成する合成手順とをコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体が提供される。

【0024】本発明のさらに他の観点によれば、手話画像のデータベースを作成するデータベース作成手順と、

10

20

30

40

50

主たる動画像を符号化して前記データベース内の手話画像と多重化する多重化手順と、前記多重化された画像から主たる動画像と手話画像を復号し、合成出力する手順とをコンピュータに実行させるためのプログラムが提供される。

【0025】本発明のさらに他の観点によれば、動画像のオブジェクト・ベース符号化を行い、手話画像のデータベースを作成するプログラムであって、手話者の動画像を入力する動画像入力手順と、手話者のみをオブジェクトとして抽出し形状情報の動画像を獲得するオブジェクト抽出手順と、前記手話者の動画像を符号化する自然画像符号化手順と、前記形状情報の動画像を符号化する形状情報画像符号化手順と、前記符号化後の自然画像と形状情報画像とを多重化する多重化手順と、前記手話画像の意味をテキストデータとして入力するテキスト入力手順と、前記多重化した画像と前記テキストデータをデータベース化して蓄積するデータベース蓄積手順とをコンピュータに実行させるためのプログラムが提供される。

【0026】本発明のさらに他の観点によれば、主たる動画像信号とオーディオ信号を符号化する符号化手順と、前記オーディオ信号から音声の情報を抽出し認識する音声認識手順と、前記認識した音声データを利用し手話画像のデータベースから所望の手話画像を検索する手話画像検索手順と、前記符号化した主たる動画像信号とオーディオ信号と前記手話画像を多重化して出力する多重化手順とをコンピュータに実行させるためのプログラムが提供される。

【0027】本発明のさらに他の観点によれば、入力信号を、主たる動画像信号及びオーディオ信号の組みと、手話画像とに分配する第1の分配手順と、前記主たる動画像信号及びオーディオ信号を復号する第1の復号手順と、前記手話画像から、手話者の自然画像と手話者の形状情報画像とに分配する第2の分配手順と、前記手話者の自然画像を復号する第2の復号手順と、前記手話者の形状情報画像を復号する第3の復号手順と、前記主たる動画像信号と前記手話者の自然画像と形状情報画像から、主たる動画像と手話画像を合成する合成手順とをコンピュータに実行させるためのプログラムが提供される。

【0028】本発明はこのように構成したので、任意形状の手話画像のデータベースを効率的に構築し、主たる動画像及びオーディオ信号（全画面の動画像と音声）に手話画像を合成することで、聴覚障害者にとって自然でわかりやすい映像サービスを充足させることができる。

【0029】

【発明の実施の形態】以下、本発明の実施形態を、図面を用いて詳細に説明する。

（第1の実施形態）図7及び図8を用いてオブジェクト・ベース符号化方法を説明する。図7(a)は自然画像、

図7(b)は図7(a)に対応した形状情報画像である。このようにオブジェクト・ベース符号化方法は、画像全体ではなくオブジェクトの部分のみを符号化することから、画像を高い効率で符号化することができる。

【0030】また、図8(a)では、主たる画像を表示した場合、図8(b)では前記画像に手話画像のオブジェクトを合成した状態を示す。このようにオブジェクト・ベース符号化では、既にオブジェクトの形状が明確であるため、その背景となる画像との合成が簡易的に可能となる。

【0031】図1は本発明の第1の実施形態としての手話画像のデータベースを作成するための装置構成を示すブロック図である。尚、本実施形態では、MPEG-4における実装方法を述べる。またMPEG-4に従い、自然画像による動画像をテクスチャ画像、このテクスチャに対応した符号化すべきオブジェクトの形状情報による動画像をシェイプ画像と称する。本装置には、前記テクスチャ画像と、前記シェイプ画像とで構成する2種類の動画像を入力する。

【0032】図1の構成を説明する。100は手話者を撮影するカメラである。101はカメラ100で撮影した手話者の画像から手話者の領域と背景領域を分割し、手話者の領域を1、背景の領域を0としたバイナリー・アルファ・プレーンで構成した動画像を生成するシェイプ生成器である。102はカメラからの手話者の自然画像を符号化するテクスチャ符号化器である。103はシェイプ生成器101からの手話者のシェイプ画像を符号化するシェイプ符号化器である。104はテクスチャ符号化器とシェイプ符号化器からのストリームを多重化する多重化器である。105は多重化器からの出力と別途入力された手話画像に対応したテキストを関連付け蓄積する手話画像データベースである。

【0033】次に図1の動作について説明する。まず、カメラ100により手話者の画像を撮影する。撮影については後に他の画像と合成しやすくするため、ブルーバックなどの背景を用いる。撮影した画像からシェイプ生成器101により手話者の形状情報画像を生成する。次に手話者自身の撮影画像（自然画像）をテクスチャ符号化器102へ、形状情報画像をシェイプ符号化器103へ入力し、それぞれをオブジェクト符号化する。それぞれの符号化データは多重化器104で多重化され手話画像ストリームとして出力され、手話画像の意味であるテキストデータと共にデータベース105へ蓄積される。

【0034】図2は、本発明の第1の実施形態としての主たるオーディオ・ビデオと手話画像を符号化し多重化する装置の全体構成を示すブロック図である。図2の構成を説明する。200は主たるコンテンツであるビデオ信号とオーディオ信号を入力し、符号化する主AV符号化器である。201は入力したオーディオから音声を認識しそのテキストデータを出力する音声認識器である。2



03は音声認識器201からのテキストデータを用いて、前記手話画像データベースから対応する手話画像を検索する手話画像検索器である。202は主AV符号化器からの主ストリームと、手話画像検索器203からの手話画像ストリームを多重化する多重化器である。

【0035】次に図2の動作について説明する。まず、主たるビデオとオーディオの素材信号は符号化器200で符号化され主ストリームとして出力される。同時にオーディオ素材信号は、音声認識器201により音声認識、解読する。その後、手話画像検索器203により解読した音声に対して蓄積された手話画像データベース105から整合する手話画像ストリームを取得する。次に多重化器202により主ストリームと手話画像ストリームを多重化して最終的なビットストリームを出力する。

【0036】図3は、本発明の第1の実施形態としての主たるオーディオ・ビデオと手話画像を復号し合成する装置の全体構成を示すブロック図である。図3の構成を説明する。300は前記ビットストリームを入力し、主ストリームと手話画像ストリームとに分配する分配器である。301は分配器300からの主ストリームをオーディオ信号とビデオ信号へ復号する主AV復号器である。302は前記手話画像ストリームから手話者のテクスチャストリームと形状情報であるシェイプストリームを分配する分配器である。303は前記テクスチャストリームから手話者のテクスチャ画像信号を復号するテクスチャ復号器である。304は前記シェイプストリームから手話者のシェイプ画像信号を復号するシェイプ復号器である。305は画像合成指示信号を入力し、画像合成指示信号がONの場合は主AV復号器301からのビデオ信号と手話画像復号器302からの信号を合成して出力し、OFFの場合は主AV復号器301からのオーディオ・ビデオ信号のみを出力する手話画像合成器である。

【0037】次に図3の動作について説明する。まず、ストリームを入力し、分配器300により主ストリームと手話画像ストリームを分配する。主ストリームは主AV復号器301により復号され主ビデオ信号と主オーディオ信号として出力される。手話画像ストリームは手話画像分配器302により手話者のテクスチャ画像信号となるテクスチャストリームと手話者のシェイプ画像信号となるシェイプストリームへ分配される。テクスチャストリームはテクスチャ復号器303により手話者のテクスチャ画像信号へ復号され、シェイプストリームはシェイプ復号器304により手話者のシェイプ画像信号へ復号される。次に主AV復号器301からの主ビデオ信号とテクスチャ復号器303からのテクスチャ画像信号とシェイプ復号器304からのシェイプ画像信号から手話画像合成器305により主ビデオ信号と手話画像信号を合成したビデオ信号を出力する。

【0038】図4にMPEG-4を用いて、主オーディオ・ビデオストリームと手話画像ストリームを多重化したとき

の全体のストリーム構成を示す。MPEG-4を用いることで、全てのストリームを一定の時間単位でパケット化して、表示時刻を示すTS(Time Stamp)を重畳後多重化することで、各ストリームの表示管理等が容易に可能となる。

【0039】図5に前記手話画像合成器305の動作フローチャートを示す。まず主オーディオ・ビデオ信号及び手話画像信号(テクスチャ画像とシェイプ画像)を入力する(400)。次に必ず出力すべき主たるビデオ信号を出力する(401)。次に手話画像を合成して表示するための指示信号を入力する(402)。合成するように指示を受けた場合、つまり合成表示指示信号がONの場合は、主ビデオに手話画像を合成したビデオ信号を出力する(404)。指示信号がOFFの場合は合成を行わず主ビデオ信号を出力し(401)。同時にオーディオ信号を出力する(403)。

【0040】第1の実施形態のその他の構成を説明する。本実施形態では、オブジェクト・ベース符号化方式について国際標準であるMPEG-4を用いているが、その他のオブジェクト・ベース符号化方式に適用してもよい。また、オブジェクト・ベース符号化方式以外の符号化及び復号はMPEG-4以外、例えばMPEG-1,2などであってもかまわない。手話画像合成器305ではオーディオと手話画像をスイッチしているが、手話画像のみを合成もしくは非出力として、オーディオは出力したままとしてもかまわない。

【0041】第1の実施形態の効果を説明する。本実施形態は、符号化した手話画像ストリームを用いて手話画像データベースを構築するため、少ないデータ容量で多くの手話画像を蓄積する手段を提供する。手話画像を自然画像にて作成するため、CGなどの画像に比較して、手話者の細やかな表情や動きを蓄積することができる。オブジェクト・ベース符号化を用いて手話画像を符号化しているため、他の画像との合成がしやすい。

【0042】(第2の実施形態)図6は、第2の実施形態におけるコンピュータの構成を示すブロック図である。500はコンピュータ全体の制御、及び種々の処理を行う中央演算装置(CPU)、501は本コンピュータの制御に必要なオペレーティングシステム(OS)、ソフトウェア、データ、演算に必要な記憶領域を提供するメモリである。また、CPU500が各種の処理を行う際のワークエリアとしても用いられる。

【0043】502は種々の装置をつなぎ、データ、制御信号をやりとりするバス。503は各種のソフトウェアを蓄積する記憶装置、504は動画データデータを蓄積する記憶装置、505は画像やコンピュータからのシステムメッセージなどを表示するモニタである。

【0044】507は通信回路に符号化データを送信する通信インターフェースであり、装置外部のLAN、公衆回線、無線回線、放送電波等と接続されている。506

はコンピュータを起動したり、ビットレート等の各種条件を設定したりするための端末である。

【0045】メモリ501にはコンピュータ全体を制御し、各種ソフトウェアを動作させるためのOSや動作させるソフトウェアを格納し、画像データを符号化のために読み込むエリア、一時的に符号データを格納する符号エリア、各種演算のパラメータ等を格納しておくワーキングエリアが存在する。

【0046】図6の動作を説明する。上記のような構成において、処理に先立ち、端末506から記憶装置504に蓄積されている手話動画像及び対するテキストデータから符号化する動画像データを選択し、コンピュータの起動が指示される。すると、記憶装置503に格納されているソフトウェアがバス502を介してメモリ501に展開され、ソフトウェアが起動される。そして、CPU500により図1、図2、図3に示した処理を順次実現する。

【0047】本実施形態におけるコンピュータは、第1の実施形態における任意形状の手話画像のデータベースを効率的に構築し、主たるAVストリーム（全画面の動画像と音声）に手話画像を合成する装置として機能する。

【0048】以上の説明から明らかなように、第1及び第2の実施形態は、オブジェクト・ベース符号化装置及びその方法を用いて、任意形状の手話画像のデータベースを効率的に構築し、主たるAVストリーム（全画面の動画像と音声）に手話画像を合成することで、聴覚障害者にとって自然でわかりやすい映像サービスを充足させる手段を提供する。

【0049】本実施形態は、コンピュータがプログラムを実行することによって実現することができる。また、プログラムをコンピュータに供給するための手段、例えばかかるプログラムを記録したCD-ROM等の記録媒体又はかかるプログラムを伝送するインターネット等の伝送媒体も本実施形態として適用することができる。上記のプログラム、記録媒体及び伝送媒体は、本発明の範疇に含まれる。記録媒体としては、例えばフレキシブルディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、磁気テープ、不揮発性のメモリカード、ROM等を用いることができる。

【0050】なお、上記実施形態は、何れも本発明を實施するにあたっての具体化の例を示したものに過ぎず、これらによって本発明の技術的範囲が限定的に解釈されてはならないものである。すなわち、本発明はその技術思想、またはその主要な特徴から逸脱することなく、様々な形で実施することができる。

【0051】

【発明の効果】以上説明したように、任意形状の手話画

像のデータベースを効率的に構築し、主たる動画像及びオーディオ信号（全画面の動画像と音声）に手話画像を合成することで、聴覚障害者にとって自然でわかりやすい映像サービスを充足させることができる。

【図面の簡単な説明】

【図1】本発明の第1の実施形態としての手話画像データベース作成装置の全体構成を示すブロック図である。

【図2】本発明の第1の実施形態としての主たるオーディオ・ビデオと手話画像を符号化し多重化する装置の全体構成を示すブロック図である。

【図3】本発明の第1の実施形態としての主たるオーディオ・ビデオと手話画像を復号し合成する装置の全体構成を示すブロック図である。

【図4】主たるオーディオ・ビデオストリームと手話画像ストリームを多重化したときの全体のストリーム構成を示す図である。

【図5】手話画像を合成する際の動作フローチャートである。

【図6】本発明の第2の実施形態としてのコンピュータの構成を示すブロック図である。

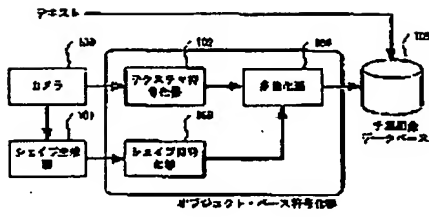
【図7】手話者をオブジェクト符号化する際の自然画像と形状情報画像の例を示す図である。

【図8】主たる画像に対して手話画像を合成した際の画像例を示す図である。

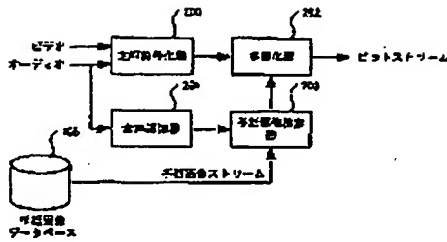
【符号の説明】

- 100 カメラ
- 101 シェイプ生成器
- 102 テクスチャ符号化器
- 103 シェイプ符号化器
- 104、202 多重化器
- 105 手話画像データベース
- 200 主AV符号化器
- 201 音声認識器
- 203 手話画像検索器
- 300 分配器
- 301 主AV復号器
- 302 手話画像分配器
- 303 テクスチャ復号器
- 304 シェイプ復号器
- 305 手話画像合成器
- 500 CPU
- 501 メモリ
- 502 バス
- 503、504 記憶メディア
- 505 モニタ
- 506 端末
- 507 通信I/F

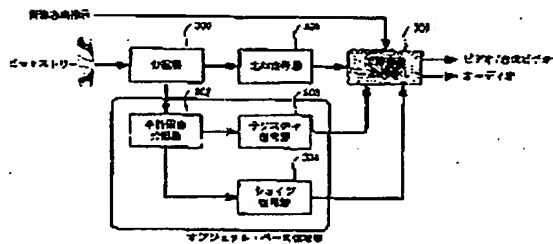
【図1】



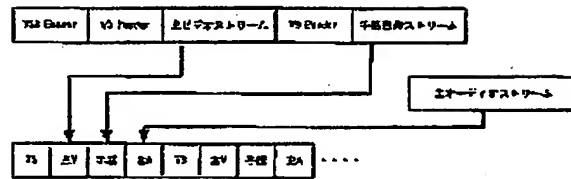
【図2】



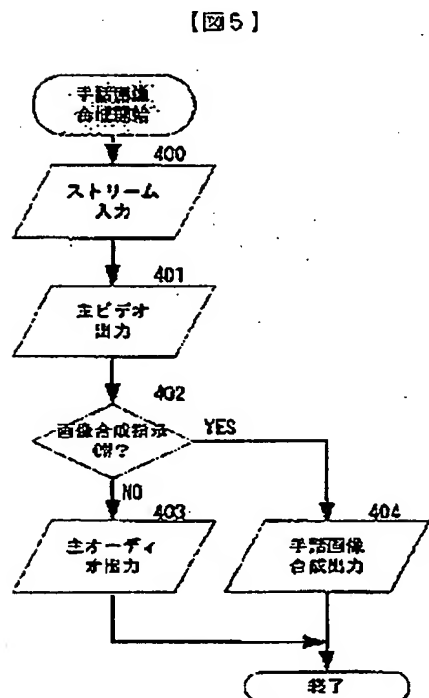
【図3】



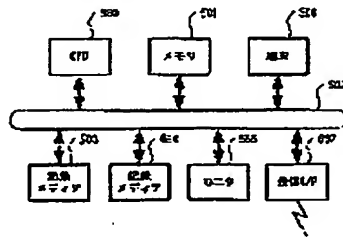
【図4】



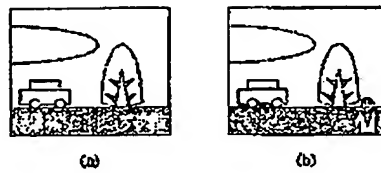
【図5】



【図6】



【図8】



フロントページの続き

(51)Int.Cl.

識別記号

F I

ターム(参考)

H04N 7/08

G10L 3/00

551G

7/081

551C

F ターム(参考) SC023 AA06 AA17 AA37 AA38 BA02  
 BA11 CA01 CA05 DA04  
 SC052 AA01 AC08 CC11 DD04 DD06  
 SC059 KK37 MA00 MB03 MB06 MB12  
 MB23 PP04 PP28 PP29 RB01  
 RB18 RC19 RC32 SS06 SS12  
 SS19 SS30 UA02 UA05  
 SC063 AB07 AC01 AC05 CA11 CA20  
 CA36  
 SD015 KK01